# A New Gene-Based test of Association Using Extended Rasch Models

Wenjia Wang, Mickal Guedj

Pharnext,Department of bioinformatics and biostatistics,
92130 Issy-Les-Moulineaux, France

**Abstract.** In GWAS analysis, gene-based tests of association has become an important alternative to the tradition single-marker association analysis. However, several statistical issues limited the performance of gene-based tests when assessed to real data. Here we introduce a new test to provide a $p$ value for a gene by using extended Rasch Models. It provide a score for each individual in GWAS by aggregate genotypes of SNPs within a gene and the weight of each SNP. In a variate of simulation, this test maintained a correct false positive rate and its power exceeded other tests when the number of disease related SNP increased. This test can be generized to multivariate traits analysis.

**Keywords:** GWAS, gene-bases association test, Rasch Models

## 1 Introduction

Genome-wide association studies (GWAS) are increasingly used for identification of genetic associations for complex diseases. Traditionally, single nucleotide polymorphism (SNP) are tested individually. Recently, a gene-based approach consider association between diseases and all SNPs within a gene becomes increasingly important. As a matter of fact, in Genetics, the gene is often considered as the unit of interest as the analyses of the functional mechanisms of a disease are generally based on genes and their products such as RNA or resulting proteins[1]. To derive a gene-level measure of significance, such as a test statistic or a p-value, one needs to combine the results of all the SNPs corresponding to the gene.

Computing a single p-value per gene raises several statistical issues. First, several SNPs are usually genotyped within a gene and combining the results of each individual SNP test outcome corresponds to a multiple-testing situation. In addition, markers within a gene are usually closely located on the genome and therefore likely to be in linkage disequilibrium (LD). This LD pattern of a gene leads to a situation of multiple-testing with dependent tests. Hence, a gene-based association approach considering this two issues is often of interest.

A number of gene-based tests have been proposed such as GATES[2], margin and VEGAS[3]. However, most current gene-based tests consider combining single SNP association $p$ value to compute a gene-based test statistic. These methods may ignores the potential joint effect of SNPs within a gene. Moreover, the permutation process to account for confounding factors required in several approaches is quite time and computation consuming.

Therefore we proposed a new gene-based association test using extended Rasch Models[4] . This test can combine the genotypes of all SNPs within a gene to produce a score for each individuals and then to derive a gene-level $p$ value.It is also less time consuming than permutation. Rasch Model is a generalized linear model for analyzing categories data to measure variables. Rasch models are increasingly being use in many areas such as education and clinics [5], but never applied to genetics. After study, we found that Rasch models can be adapted to the analysis of GWAS data. GWAS data is consisted by a set of SNP that values are categorized in 0, 1 and 2 as items with 3 categories. The probability of linkage between a set of SNP and disease can be estimated as latent trait in the Rasch models, so a p-value can be derived for each gene. Compared to other statistical tests which calculate p-value for each SNP, Rasch models consider the pattern of every SNP in a gene. The application of Rasch on GWAS data may offer a better solution for genetic disorders research.

In a series of simulation with different scenarios (number of DSL, relative risk, LD structure of genes) we compared the extended Rasch Models to 6 other gene-based association tests: minP[6], margin test[7], goeman test[8], GATES, SKAT, Fisher's method. In the comparison, the extented Rasch models maintain correct false positive rate in different situation and it has the highest power when the number of disease related SNPs exceeds 7 in simulation.

This gene-based test can be further extended to treat GWAS data with multivariate phenotypes, which is of interest in GWAS study but few approaches have been developed for.

## References

1. Eric Jorgenson and John S Witte. A gene-centric approach to genome-wide association studies. *Nature Reviews Genetics*, 7(11):885–891, 2006.
2. Miao-Xin Li, Hong-Sheng Gui, Johnny SH Kwan, and Pak C Sham. Gates: a rapid and powerful gene-based association test using extended simes procedure. *The American Journal of Human Genetics*, 88(3):283–293, 2011.
3. Jimmy Z Liu, Allan F Mcrae, Dale R Nyholt, Sarah E Medland, Naomi R Wray, Kevin M Brown, Nicholas K Hayward, Grant W Montgomery, Peter M Visscher, Nicholas G Martin, et al. A versatile gene-based test for genome-wide association studies. *The American Journal of Human Genetics*, 87(1):139–145, 2010.
4. Georg Rasch. Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests. 1960.
5. Jeremy C Hobart, Stefan J Cano, John P Zajicek, and Alan J Thompson. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *The Lancet Neurology*, 6(12):1094–1105, 2007.
6. Ali Torkamani, Eric J Topol, and Nicholas J Schork. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics*, 92(5):265–272, 2008.
7. Wei Pan. Asymptotic tests of association with multiple snps in linkage disequilibrium. *Genetic epidemiology*, 33(6):497–507, 2009.
8. Jelle J Goeman, Sara A Van De Geer, and Hans C Van Houwelingen. Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):477–493, 2006.